Microsoft Word Author Guidelines for CVPR Proceedings

Anonymous CVPR submission  
  
Paper ID \*\*\*\*< replace \*\*\*\* here and in header with paperID>

**Abstract**

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word “Abstract” as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. The abstract may be up to 3 inches (7.62 cm) long. Leave two blank lines after the Abstract, then begin the main text.

# Introduction

This project’s primary was to explore the intersection of deep learning and reinforcement learning. Specifically, I focused on first implementing and then attempting to improve upon versions of work described the work first described in the original Deep Q-Network paper by Mnih et al. [1] and the follow-up paper on Double Deep Q-Networks by Hasselt et al. [2]. This was accomplished by building and training multiple types of CNNs and replay buffers to play the game *Breakout* for the Atari 2600. Everything was built making use of Gymnasium, a Python reinforcement learning framework that provides a standardized API to access the Arcade Learning Environment (ALE), itself a low-level emulator for Atari 2600 games written in C++. I was successful in creating game playing agents with a moderate, though significantly above “naïve”, degree of skill. They are likely not as proficient as the agents produced by the original researchers, primarily due to compute and training time constraints, though a 1:1 comparison is difficult due to variations in training and testing methodologies. I was, however, able to meaningfully exceed the performance benchmarks defined as “success” in the Univ. of Illinois homework assignment [3] that Prof. Kosecka referred me to as a starting point.

# Theoretical Background

For the standard DQN and DDQN I relied on the theoretical underpinnings introduced in the original papers, but for context a brief summarization is provided in this section. *Breakou*t, like any Atari game, consists of a series of frames where each image displayed on the screen, is accompanied by an action . This combination could be considered a state in an of itself, but it lacks important temporal context. Thus, I instead define a state as a series (of history length, of frames and actions, such that A sequence of these states that terminates at a finite point makes up a Markov decision process (MDP), to which we can apply the well-known *Bellman equation*. This defines the value of taking a particular action at given state as , i.e. the Q-value of an action is the reward from the action plus the expected discounted value of the value maximizing action taken at the next state, and so on. Using traditional dynamic programming reinforcement learning algorithms (typically Value Iteration or Policy Iteration) this can be solved analytically. However, for even moderately complex games with large numbers of states and/or actions, this is practically and/or computationally intractable. This provides the motivation to instead use deep learning techniques to attempt to directly “learn” the Q-function by inputting (state, action) pairs into a neural network that outputs the Q-values for the set of actions that can be taken at that state.

## Experience Replay

An obvious hurdle to implementing this as a training task is the lack of a readily identifiable “ground-truth” to compare learned Q-values against. If the correct Q-values for a state in a game were known, the game would already have a solved perfect strategy. Hence, deep reinforcement learning is inherently an unsupervised learning task. The solution is to use the neural network itself to itself to generate the target to train against. Prior to the Mnih paper, my understanding is the typical approach was to use a network (the *target* network) with weight parameters updated in the prior iteration () to compute the target, . The loss can then be calculated as

, where . One of key insights of the Mnih paper was to instead build an Experience Replay , a buffer of states collected during training as the agent was playing using an ε-greedy strategy (defined as taking a random action ε percent of the time and using the target network to find the max Q-value action (1- ε) percent of the time.

## DQN vs Double DQN

The standard DQN process (which I followed in my implementation) relies on sampling a random batch of states and from the replay buffer, passing the states to the current iteration of the network (the *policy* network) and the states to the target network, then calculating a loss and performing stochastic gradient descent to update the policy network. The target network parameters are then updated to match the policy network periodically (the exact frequency is controlled via a hyperparameter typically in the range of once every 1K – 10K training steps)

While this works quite well, but it does have the downside of creating positive feedback loops that lead to overestimated Q-values, due to both the action selection and value estimation being done using the same Q-function. The algorithm selects the maximum Q-value from noisy estimates, tending to select overestimated values which creates a bias that is then propagated into the learning target as well, causing the network to learn even higher Q-values.

To address this issue, the Hasselt et al. paper introduced the concept of a Double DQN. It still makes use of both a policy network and target network, but the key difference is that while actions are selected based on the Q-values calculated using the policy network, the training target is calculated using Q-values from the target network for that policy network selected action. The paper showed that this is successful in reducing overestimation bias (the mathematics of which are beyond the scope of this writeup), and it generally led to more successful agents (i.e. produced higher scores) across a variety of Atari games. Notably however, *Breakout* was one of a minority of games where the Hasselt paper actually reported higher results for the standard DQN – something I saw as well in my empirical results (to be discussed further later).

## The ruler

The LATEX style defines a printed ruler which should be present in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document using a non-LATEX document preparation system, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera ready copy should not contain a ruler. (LATEX users may uncomment the \cvprfinalcopy command in the document preamble.) Reviewers: note that the ruler measurements do not align well with lines in the paper—this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. Just use fractional references (e.g. this line is 189.5), although in most cases one would expect that the approximate location will be adequate.

## Mathematics

Please number all of your sections and displayed equations. It is important for readers to be able to refer to any particular equation. Just because you didn’t refer to it in the text doesn’t mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like “the equation second from the top of page 3 column 1”. (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin’s description of how to write mathematics: <http://www.cvpr.org/doc/mermin.pdf>.

Sdflklj

## Blind review

Figure : Example of a figure with caption. Captions are set in roman, 9 point. Use a Drawing area to make space for figures.

Many authors misunderstand the concept of anonymizing for blind review. Blind review does not mean that one must remove citations to one’s own work—in fact it is often impossible to review a paper unless the previous citations are known and available.

Blind review means that you do not use the words “my” or “our” when citing previous work. That is all. (But see below for techreports)

Saying “this builds on the work of Lucy Smith [1]” does not say that you are Lucy Smith, it says that you are building on her work. If you are Smith and Jones, do not say “as we show in [7]”, say “as Smith and Jones show in [7]” and at the end of the paper, include reference 7 as you would any other cited work.

An example of a bad paper:

An analysis of the frobnicatable foo filter.

In this paper we present a performance analysis

of our previous paper [1], and show it to be inferior

to all previously known methods. Why the

previous paper was accepted without this analysis

is beyond me.

[1] Removed for blind review

An example of an excellent paper:

An analysis of the frobnicatable foo filter.

In this paper we present a performance analysis of

the paper of Smith et al. [1], and show it to be

inferior to all previously known methods. Why the

previous paper was accepted without this analysis

is beyond me.

[1] Smith, L and Jones, C. “The frobnicatable

foo filter, a fundamental contribution to human

knowledge”. Nature 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [4] as additional material and cite it as

[1] Authors. “The frobnicatable foo filter”, ECCV 2006 Submission ID 324, Supplied as additional material eccv06.pdf.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not require the reviewer to go to a techreport for further details. Thus, you may say in the body of the paper “further details may be found in [5]”. Then submit the techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let’s say it’s 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled ”Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties”, by Zeus *et al.*

You can handle this paper like any other. Don’t write “We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]”. That would be silly, and would immediately identify the authors. Instead write the following:

We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] didn’t handle case B properly. Ours handles it by including a foo term in the bar integral.

...

The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don’t you know. It displayed the following behaviours which show how well we solved cases A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus et al, but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

## Miscellaneous

When citing a multi-author paper, you may save space by using “*et alia*”, shortened to “*et al*.” (not “*et. al.*” as “*et*” is a complete word.) However, use it only when there are three or more authors. Thus, the following is correct:

“Frobnication has been trendy lately. It was introduced

by Alpher [3], and subsequently developed by Alpher and Fotheringham-Smythe [1], and Alpher *et al.* [2].”

This is incorrect: “... subsequently developed by Alpher et al. [1] ...” because reference [1] has just two authors. If you use the \etal macro provided, then you need not worry about double periods when used at the end of a sentence as in Alpher et al.

For this citation style, keep multiple citations in numerical (not chronological) order, so prefer [1, 3, 4] to [3, 1, 4].

# Formatting your paper

All text must be in a two-column format. The total allowable width of the text area is inches (17.5 cm) wide byinches (22.54 cm) high. Columns are to be 31/4 inches (8.25 cm) wide, with a 5/16 inch (0.8 cm) space between them. The main title (on the first page) should begin 1.0 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be inches (2.86 cm) from the bottom edge of the page for 8.5 × 11-inch paper; for A4 paper, approximatelyinches (4.13 cm) from the bottom edge of the page.

## Margins and page numbering

Figure 0: Short captions should be centred.

All printed material, including text, illustrations, and charts, must be kept within a print area  inches (17.5 cm) wide by inches (22.54 cm) high.

## Type-style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title 1-3/8 inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The ABSTRACT and MAIN TEXT are to be in a twocolumn format.

MAIN TEXT. Type main text in 10-point Times, singlespaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. 1/6 inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figures 1 and 2. Short captions should be centred. Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

FIRST-ORDER HEADINGS. (For example, 1. Introduction) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. Should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

## Footnotes

Please use footnotes[[1]](#footnote-1) sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

## References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [4]. Where appropriate, include the name(s) of editors of referenced books.

## Illustrations, graphs, and photographs

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Many readers (and reviewers), even of an electronic copy, will choose to print your paper in order to read it.

You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic. When placing figures in LATEX, it’s almost always best to use \includegraphics, and to specify the figure width as a multiple of the line width as in the example below

\usepackage[dvips]{graphicx} ...

\includegraphics[width=0.8\linewidth]

{myfile.eps}

## Color

Color is valuable, and will be visible to readers of the electronic copy. However ensure that, when printed on a monochrome printer, no important information is lost by the conversion to grayscale.

# Final copy

You must include your signed IEEE copyright release form when you submit your finished paper. We MUST have this form before your paper can be published in the proceedings. Please direct any questions to the production editor in charge of these proceedings at the IEEE Computer Society Press: Phone (714) 821-8380, or Fax (714) 761-1784.

# References

1. A. Alpher, , and J. P. N. Fotheringham-Smythe. Frobnication revisited. Journal of Foo, 13(1):234–778, 2003.
2. A. Alpher, , J. P. N. Fotheringham-Smythe, and G. Gamow. Can a machine frobnicate? Journal of Foo, 14(1):234–778, 2004.
3. A. Alpher. Frobnication. Journal of Foo, 12(1):234–778, 2002.
4. Authors. The frobnicatable foo filter, 2006. ECCV06 submission ID 324. Supplied as additional material eccv06.pdf.
5. Authors. Frobnication tutorial, 2006. Supplied as additional material tr.pdf.

1. This is what a footnote looks like. It often distracts the reader from the main flow of the argument. [↑](#footnote-ref-1)